Distributed-Parameter Large Basin Runoff Model. I: Model Development

Thomas E. Croley II[1] and Chansheng He[2]

Abstract: We present a case study of modifying an existing macroscale rainfall-runoff model, the large basin runoff model (LBRM), developed at NOAA's Great Lakes Environmental Research Laboratory, to the microscale in a two-dimensional representation. First, we review the LBRM and then describe changes in several process submodels, which were originally designed specifically for large areas. We also change the model structure so that we may use the LBRM on an individual cell at the microscale within a watershed. We then discuss spatial scaling of model parameters to enable an (initial) application to the microscale with parameters available from the macroscale. We then organize watershed cells and flow routing and conclude with notes on computer implementation. In the accompanying companion paper, we present details of the model calibration, application, and experimentation on the Kalamazoo River watershed.

DOI: 10.1061/(ASCE)1084-0699(2005)10:3(173)

CE Database subject headings: Hydrology; Watersheds; Spatial distribution; Flow distribution; Parameters; Runoff.

Background

Climate-change research and the increasing demand for effective water resources management, together with rapid advances in the ready availability of satellite imagery and other multiple databases and computing technology, have led to a proliferation of hydrologic models. Many were developed for research at the microscale and are designed (particularly distributed process models) to simulate hydrological processes in great detail for small geographic areas (usually $<10^2$ km^2). They describe important rainfall runoff components and processes with differential equations, e.g., the Système Hydrologique Europèen model and THALES model (Beven 2000) and require multiple databases to compute spatial and temporal distributions of energy and water balances in the soil-plant-atmosphere system (Beven 2000). Such detailed input data are often expensive and time-consuming to collect for a large watershed (say, 10^3 to 10^4 km^2). Since a large watershed may be discretized into thousands of grid cells, this type of model requires much computational power, challenging even current computational technology. Also, with many different parameters involved for each element, parameter calibration becomes extremely difficult.

Macroscale operational hydrologic models, unlike microscale watershed models, are defined over large areas ($>10^3$ km^2) and long timescales, typically for use over monthly and annual or longer timescales at a daily interval. Large-scale operational hydrology models serve as a linkage between the research community and policy/decision support institutions. Such models describe important hydrological processes as a system of interconnected storages that are recharged and depleted in accordance with mass continuity equations. They use a set of distribution functions—statistical, simple functional forms, unit hydrographs, hydrological response units, or hydrological indices—to represent the spatial variability of runoff generation, rather than full process descriptions, across the study watershed (Beven 2000). To support sustainable water resources applications over large areas, macroscale models must have few parameters to calibrate, use easily accessible meteorological and hydrologic databases, and be user-friendly. Indeed, compared with microscale models, macroscale models are driven by the availability of existing databases and usually require fewer parameters (Mohseni and Stefan 1998). They are often constrained by limited data availability, computational requirements, and model application costs over larger areas. Model parameters or parameter functions are calibrated against the observed runoff data at the watershed outlet and remain the same for the entire study area. Such models include the Stanford Watershed Model (Crawford and Linsley 1966), the Xiananjian Model (Zhao et al. 1980), the United States Geological Survey's Precipitation-Runoff Modeling System (Leavesley and Stannard 1995), the Hydrologic Simulation Program in *FORTRAN* (Bicknell et al. 1996), and the Large Basin Runoff Model (Croley 2002).

Accurate accounting of soil water storage has a dominant influence on watershed runoff modeling. However, frequent spatial measurements of soils (or soil water content) are not currently available on a routine basis (Engman and Gurney 1991). Researchers often use either soil maps or such databases as STATSGO to extract soil moisture and soil characteristics for hydrologic models (Liang et al. 1996; Zhu and Mackay 2001) or to estimate soil moisture storage through calibration (Croley 2002). Alternatively, microwave remote sensing is promising for higher spatial and temporal resolutions (Engman and Gurney 1991).

Model components should include land surface, soil zones, and groundwater to produce realistic estimates of water fluxes (Koren et al. 1999; Martinez et al. 2001). Water budget computations are very sensitive to the number of layers modeled in the

[1]Research Hydrologist, Great Lakes Environmental Research Laboratory, 2205 Commonwealth Blvd., Ann Arbor, MI 48105-2945.

[2]Associate Professor, Dept. of Geography, Western Michigan Univ., 3234 Wood Hall, Kalamazoo, MI 49008-5424.

Note. Discussion open until October 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on July 7, 2003; approved on September 7, 2004. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 10, No. 3, May 1, 2005. ©ASCE, ISSN 1084-0699/2005/3-173–181/$25.00.

soil profile under wet conditions, and an insufficient number of soil layers can lead to large errors in modeled water fluxes. For modeling soil water storage, a single layer in both the upper and bottom soil zones is adequate (Martinez et al. 2001). Models employing variable source area concepts (runoff from, and infiltration into, a dynamically changing surface area) produce more accurate overland flow estimates than models using the Hortonian infiltration capacity concept (Quinn et al. 1995; Abdulla et al. 1996; Valeo and Moin 2001).

Microscale models cannot be directly transferred to large scales for water resource applications because of the nonlinearity of hydrologic responses (caused by the effect of antecedent conditions and the change of flow velocity with discharge) at different scales. Neither can macroscale models be directly applied to microscales for hydrologic simulations. A scaling function, that is, the appropriate application of information gathered at one scale to other scales, has to be identified and used to apply microscale models to large areas (upscaling) or macroscale models to small areas (downscaling). Although studies have been done on scaling, it is still an evolving topic. Since the macroscale operational models serve as a linkage between the research community and policy and decision support institutions, identifying the needs and challenges of large-scale operational models is important for better support of effective water resources decision making.

This paper and its companion (Croley et al. 2005), through a case study, address issues involved in the application of an existing large-scale hydrologic model, developed at the macroscale, to the microscale in a two-dimensional (2D) representation. This paper first reviews the large basin runoff model (LBRM), developed at the Great Lakes Environmental Research Laboratory (GLERL) for Great Lakes applications. Then it describes changes in several process submodels, which were originally specifically designed for large areas, and changes in model structure necessary for use on an individual cell at the microscale. It then discusses spatial scaling of a study watershed, organizes watershed cells and flow routing, and concludes with notes on computer implementation. The companion paper presents details of application and experimentation on the Kalamazoo River watershed, first discussing the assemblage of distributed data for the area and the calibration of the modified 2D LBRM. It then applies the model to the microscale in a first approximation and presents application results. They include several experiments in spatial parameter estimation and calibration. The companion paper finally discusses alternative model structures for future development and improvement and necessary next steps.

Large Basin Runoff Model Structural Modification

GLERL developed a large-scale operational model in the 1980s for estimating rainfall/runoff relationships on the 121 large watersheds surrounding the Laurentian Great Lakes. It is physically based to provide good representations of hydrologic processes and to ensure that results are tractable and explainable. It is used here in application to individual subareas within a watershed by modifying its structure to accept upstream flow.

The amended mass-balance schematic, written for a small piece of the watershed (cell), is shown in Fig. 1; it is the same conceptual LBRM schematic employed by Croley (2002), except for the addition of the upstream surface flow component from the upstream cell, h. Precipitation enters the snowpack, if present, and is then available as snowmelt, depending mainly on air temperature and solar radiation. Snowmelt and rainfall partly infiltrate into the soil and partly run off directly to surface storage, depending upon the moisture content of the soil. Infiltration is high if the soil is dry, and surface runoff is high if the soil is saturated. Soil moisture evaporates or is transpired by vegetation, depending on the types of vegetation, the season, solar radiation, air temperature, humidity, and wind speed. The remainder percolates into deeper basin storages that feed surface storage through interflows and groundwater flows. Generally, these supplies are high if the soil and groundwater storages are large. Finally, there is a flow into surface storage from the upstream cell, which is routed, along with all the other flows into surface storage, through the cell into the next downstream cell.

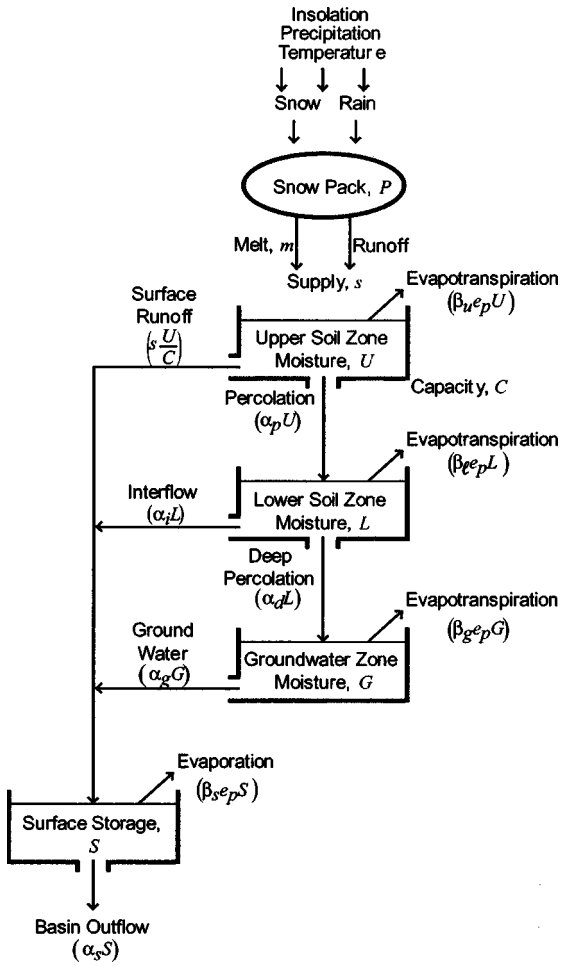

Fig. 1. Tank cascade schematic

In Fig. 1, daily precipitation, temperature, and insolation (the latter available from meteorological summaries as a function of location) may be used to determine snowpack accumulations and net supply, s. The net supply is divided into surface runoff, sU/C, and infiltration to the upper soil zone, $s-(sU/C)$, in relation to the upper soil-zone moisture content, U, and the fraction it represents of the upper soil-zone capacity, C. Percolation to the lower soil zone, $\alpha_p U$, and evapotranspiration, $\beta_u e_p U$, are taken as outflows from a linear reservoir (flow is proportional to storage). Likewise, interflow from the lower soil zone to the surface, $\alpha_i L$; evapotranspiration, $\beta_\ell e_p L$; and deep percolation to the groundwater zone, $\alpha_d L$, are linearly proportional to the lower soil-zone moisture content, L. Groundwater flow, $\alpha_g G$, and evapotranspiration from the groundwater zone, $\beta_g e_p G$, are linearly proportional to the groundwater-zone moisture content, G. Finally, basin out-

flow, $\alpha_s S$, and evaporation from the surface storage, $\beta_s e_p S$, depend on its content, S. Additionally, evaporation and evapotranspiration are dependent on potential evapotranspiration, e_p, as determined by joint consideration of the available moisture and the heat balance over the watershed. The alpha coefficients (α) are used to represent linear reservoir proportionality factors, and the beta coefficients (β) are used to represent partial linear reservoir coefficients associated with evapotranspiration.

Mass conservation equations, presented by Croley (2002) for the case of no upstream cell flow ($h=0$), are repeated here for convenience as differential equations with respect to time t.

$$\frac{d}{dt}U = s\left(1 - \frac{U}{C}\right) - \alpha_p U - \beta_u e_p U \tag{1}$$

$$\frac{d}{dt}L = \alpha_p U - \alpha_i L - \alpha_d L - \beta_\ell e_p L \tag{2}$$

$$\frac{d}{dt}G = \alpha_d L - \alpha_g G - \beta_g e_p G \tag{3}$$

$$\frac{d}{dt}S = s\frac{U}{C} + \alpha_i L + \alpha_g G - \alpha_s S - \beta_s e_p S \tag{4}$$

In each case, mass continuity yields a first-order linear differential equation of the general form

$$dZ + (\sum \alpha) Z dt = g(t)dt \tag{5}$$

where Z=storage; ($\Sigma\alpha$)=sum of linear reservoir constants for all outflows; and $g(t)$=sum of time-dependent inflows. Standard procedures (Rainville 1964) yield the general solution

$$Z_t = e^{-(\Sigma\alpha)t}\left[Z_0 + \int_0^t g(u)e^{(\Sigma\alpha)u}du\right] \tag{6}$$

where the subscript is time. Since data on precipitation and temperature are available only in time increments of a day or larger, the solutions to Eqs. (1)–(4) assume that net supply and potential evapotranspiration are distributed uniformly over the time increment. Storage values at the end of a time increment are computed from values at the beginning. In the analytical solution, results from one storage zone are used in other zones, where their outputs appear as inputs. There are several different solutions, depending upon the relative magnitudes of all coefficients in Eqs. (1)–(4). Croley (2002) solved the equations, yielding storages at the end of a time increment (U_t, L_t, G_t, and S_t) as functions of the inputs, parameters, and beginning-of-time-increment storages (storages at the end of the previous time increment: U_0, L_0, G_0, and S_0). Since the variables s and e_p change from one time increment to another, the appropriate analytical result, as well as its solution, varies with time. Mathematical continuity among solutions is preserved, however. These results are summarized elsewhere (Croley 1982). The general solution for surface storage from Eq. (4) via Eqs. (5) and (6) is

$$S_t = e^{-(\alpha_s+\beta_s e_p)t}\left[S_0 + \int_0^t \left(s\frac{U}{C} + \alpha_i L + \alpha_g G\right)e^{(\alpha_s+\beta_s e_p)u}du\right] \tag{7}$$

Croley (2002) shows that, in all cases, flow volumes are determined directly, since outflow volumes are related by their ratio of linear reservoir coefficients. In particular, the volume of basin outflow (from the surface storage) over the time increment $(0,t)$, Q, is

$$Q = (V_r + V_i + V_g + S_0 - S_t)\frac{\alpha_s}{\alpha_s + \beta_s e_p} \tag{8}$$

where V_r=surface runoff volume; V_i=interflow volume; and V_g=groundwater flow volume, all into the surface storage over the time increment $(0,t)$.

By using the above nomenclature for the case with no upstream cell flow to the surface storage ($h=0$) and a "prime" notation for the case with upstream cell flow, Eq. (4) becomes

$$\frac{d}{dt}S' = s\frac{U}{C} + \alpha_i L + \alpha_g G - \alpha_s S' - \beta_s e_p S' + h \tag{9}$$

Solving Eq. (9) via Eqs. (5) and (6) yields

$$S'_t = e^{-(\alpha_s+\beta_s e_p)t}\left[S'_0 + \int_0^t \left(s\frac{U}{C} + \alpha_i L + \alpha_g G + h\right)e^{(\alpha_s+\beta_s e_p)u}du\right] \tag{10}$$

If we approximate h as constant over the time increment $(0,t)$, Eq. (10) becomes

$$S'_t = e^{-(\alpha_s+\beta_s e_p)t}\left[S'_0 + \int_0^t \left(s\frac{U}{C} + \alpha_i L + \alpha_g G\right)e^{(\alpha_s+\beta_s e_p)u}du\right] + e^{-(\alpha_s+\beta_s e_p)t}h\int_0^t e^{(\alpha_s+\beta_s e_p)u}du \tag{11}$$

From Eqs. (7) and (10), at $t=0$, $S'_0=S_0$; therefore, Eq. (11) becomes

$$S'_t = S_t + h\frac{[1 - e^{-(\alpha_s+\beta_s e_p)t}]}{\alpha_s + \beta_s e_p} \tag{12}$$

Likewise, Eq. (8) becomes

$$Q' = (V_r + V_f + V_g + S'_0 - S'_t + ht)\frac{\alpha_s}{\alpha_s + \beta_s e_p} = Q + h\left[t - \frac{(1 - e^{-(\alpha_s+\beta_s e_p)t})}{\alpha_s + \beta_s e_p}\right]\frac{\alpha_s}{\alpha_s + \beta_s e_p} \tag{13}$$

Therefore, the output of the LBRM, applied to a single cell with no inflow from an upstream cell, (S_t and Q) can be corrected each time increment with Eqs. (12) and (13) to reflect the presence of an inflow, h, from an upstream cell (S'_t and Q'). The beginning storage in the following time increment is set equal to the ending storage for the present time increment, S'_t. The outflow volume from the cell, Q', determines the inflow to the next downstream cell; again approximating it as constant over the time interval, it is determined by dividing by the length of the time interval.

Evapotranspiration

At any instant, the evapotranspiration rate is proportional to the amount of water available as in Eqs. (1)–(4), reflecting both areal coverage and extent of supply, and to the rate of nonlatent heat released to the atmosphere (atmospheric heating), dH/dt (Croley 2002), called potential evapotranspiration, e_p

$$e_p = \frac{dH}{dt} \Big/ (\rho_w \gamma_v) \tag{14}$$

where γ_v=latent heat of vaporization; ρ_w=density of water; and H=nonlatent heat released to the atmosphere during the day.

Complementary Actual and Potential Evapotranspiration

The LBRM considers that part of the total available heat in a day is used in evapotranspiration, and the rest of it determines the potential evapotranspiration; that is, potential and actual evapotranspiration are complementary. This concept is an appropriate one for modeling very large watersheds that interact with the overlying atmosphere.

$$\Psi = H + \rho_w \gamma_v (E_u + E_\ell + E_g + E_s) \tag{15}$$

where Ψ=total heat available for evapotranspiration during the day; and where E_u, E_ℓ, E_g, and E_s=daily evapotranspiration from the upper soil zone, the lower soil zone, the groundwater zone, and the surface zone, respectively. The value of e_p is determined by simultaneous solution of Eqs. (1)–(4) and the following complementary relationship between actual evapotranspiration and that still possible from atmospheric heat, derived from Eqs. (14) and (15):

$$\int_0^d [e_p + (\beta_u U + \beta_\ell L + \beta_g G + \beta_s S) e_p] dt = \Psi / (\rho_w \gamma_v) \tag{16}$$

The evaporation from stream channels and other water surfaces (surface zone) in a large basin is very small compared with the basin evapotranspiration; groundwater evapotranspiration is also taken here as being relatively small. By taking e_p as uniform over the day and ignoring evapotranspiration from the surface and groundwater zones, Eq. (16) yields

$$e_p \cong \frac{1}{d\rho_w \gamma_v} \frac{\Psi}{1 + \beta_u \overline{U} + \beta_\ell \overline{L}} \tag{17}$$

where $\overline{U}$=average water volume in the upper soil zone over the day; $\overline{L}$=average water volume in the lower soil zone over the day; and d=time in one day. As expected, both potential and actual evapotranspiration depend on the available water supply. If the water supply is large, the daily actual evapotranspiration volume approaches the limit of the water supply, or $\Psi / \rho_w \gamma_v$, and the daily potential evapotranspiration volume approaches zero. If the water supply is small, the daily actual evapotranspiration volume approaches zero and the daily potential evapotranspiration volume approaches $\Psi / \rho_w \gamma_v$.

In the LBRM, the total heat available for evapotranspiration during the day is estimated empirically from average daily air temperature

$$\Psi = \mathbb{C} e^{T/T_b} \tag{18}$$

where T=daily average air temperature; T_b="base" temperature (a parameter to be determined in calibration); and $\mathbb{C}$=units and proportionality coefficient, determined over all days, d, from a long-term heat balance

$$\mathbb{C} = \frac{\sum_{\text{all d}} \Psi_d}{\sum_{\text{all d}} e^{T_d/T_b}} = \frac{\sum_{\text{all d}} (\Phi_d - m_d \rho_w \gamma_f)}{\sum_{\text{all d}} e^{T_d/T_b}} \tag{19}$$

where Ψ_d=total heat available on day d; Φ_d=daily surface solar insolation; m_d=daily snowmelt; and γ_f=latent heat of fusion. Note that $\mathbb{C}$ is given by Eq. (19) and not by parameter calibration.

The daily surface solar insolation, Φ_d, is calculated in two manners. The first method is present in the original LBRM; dropping the daily subscript

$$\Phi = \Phi_s \left[b_1 + b_2 \min\left(\frac{T_{\max} - T_{\min}}{15\,^\circ\text{C}}, 1 \right) \right] \tag{20}$$

where Φ_s=cloudless daily solar insolation at the surface, interpolated from midmonthly climatic values for the location (Gray 1970); b_1 and b_2=empirical coefficients relating solar insolation at the surface to cloud cover (Gray 1970); and $T_{\max}$ and $T_{\min}$=maximum and minimum daily air temperature. The difference in maximum and minimum temperatures divided by 15°C is a proxy for cloud cover (Crawford and Linsley 1966).

The second method was reverse engineered from an available weather-generation model (Richardson and Wright 1984) that provides daily values for precipitation, maximum and minimum air temperature, and solar radiation. The model accounted for persistence in each variable, the dependence among variables, and seasonal characteristics of each variable. Parameters are available for the lower 48 contiguous states. The reverse-engineered method gives daily solar insolation as a function of location, day of the year, minimum and maximum air temperatures, and precipitation for both the current and preceding day, as well as the preceding day's insolation.

Independent Actual and Potential Evapotranspiration

Evapotranspiration and potential evapotranspiration cannot be regarded as complementary when the LBRM is applied to a small cell; they must be replaced with concepts that make sense for the small scale. A more traditional independent concept, that actual evapotranspiration does not affect potential evapotranspiration, is much more appropriate for small areas, such as each of the cells used in the distributed-parameter LBRM application. Eq. (17) becomes

$$e_p \cong \frac{\Psi}{d\rho_w \gamma_v} \tag{21}$$

Spatial Scaling

Upper, Lower, and Groundwater Tanks

The linear reservoir coefficients, derived for some of the moisture storages (U, L, and G) from lumped-parameter applications of the model, should apply, at least in a first approximation, to the small scale as well as to the large scale. This statement is true because the LBRM uses linear reservoirs to model storages. We consider n identical moisture storages, placed side by side; each is represented as a linear reservoir

$$q_i = \alpha Z_i, \quad i = 1, \ldots, n \tag{22}$$

The total outflow is the same function of the total storage; we simply sum over the n storages

$$\sum_{i=1}^{n} q_i = \alpha \sum_{i=1}^{n} Z_i \tag{23}$$

Eq. (23) allows us to expect that the linear reservoir coefficients of U, L, and G for the individual cells are on the same order of magnitude as the coefficient for the entire area. This outcome provides a basis for approximating the parameters of the LBRM applied to individual cells of a watershed with the parameters of the LBRM applied to the encompassing, spatially integrated watershed. Refinements quickly come by changing the first-

approximation coefficients to reflect the perceived hydrology of each of the cells.

Surface Tank

Of course, the total flow, summed in the manner of Eq. (23), would not be the same as the flow at the outlet of the combined surface area; it would depend on how the individual flows are routed along the surface. The total flow at the outlet of the watershed represents outflow from a cascade of linear tanks. For the lumped-parameter case, one tank represents the entire watershed surface. To get an idea of how parameters are related, we consider the following simple comparison. Instead of a single linear reservoir representing the entire watershed surface (with parameter α), we consider a series of n identical linear reservoirs, each with parameter ω, where each one empties into the one below it. The response at the outlet to an initial unit volume (at time $t=0$) is (Chow 1964; Nash 1957)

$$q(t) = \frac{\omega^n t^{n-1} e^{-\omega t}}{(n-1)!} \tag{24}$$

For the case of a single linear reservoir ($n=1$ and $\omega=\alpha$)

$$q(t) = \alpha e^{-\alpha t} \tag{25}$$

By solving Eqs. (24) and (25) for a characteristic time of some kind, we can derive a relationship between α and ω to use for approximating ω. We recognize Eq. (24) as the gamma distribution; replacing the factorial in Eq. (24) with the gamma function $\Gamma(n)$ and integrating to get the accumulated outflow volume $Q(t)$ at time t

$$Q(t) = \frac{1}{\Gamma(n)} \int_0^t \omega^n u^{n-1} e^{-\omega u} du \tag{26}$$

Transforming Eq. (26) with $x=\omega u$

$$Q(x) = \frac{1}{\Gamma(n)} \int_0^x x^{n-1} e^{-x} dx \tag{27}$$

We define the characteristic time, t_γ, as the time when the fraction of runoff, γ, occurs. In terms of the transform

$$t_\gamma = \frac{x_\gamma}{\omega} \tag{28}$$

where $x_\gamma=\gamma$-quantile from Eq. (27). By taking quantiles from Eq. (27) for both the n-reservoir case and the single reservoir case ($n=1$), equating the characteristic times ($t_\gamma^n=t_\gamma^1$, where the superscripts denote the number of linear reservoirs), and solving for ω in terms of α with Eq. (28)

$$\omega = \alpha \frac{x_\gamma^n}{x_\gamma^1} \tag{29}$$

Fig. 2 contains the ratio in Eq. (29) for different values of γ and n; it is almost a linear function of n. Thus, if we have an estimate of α from a lumped-parameter application of the LBRM over a watershed and if we know the order of the cascade of cells in the application of the distributed model, we can estimate a starting value for the linear reservoir coefficient to apply to every cell's surface tank (assuming spatial uniformity), depending on which characteristic times are important to preserve. Refinements again quickly come by changing this first-approximation coefficient to reflect the perceived hydrology of each of the cells.

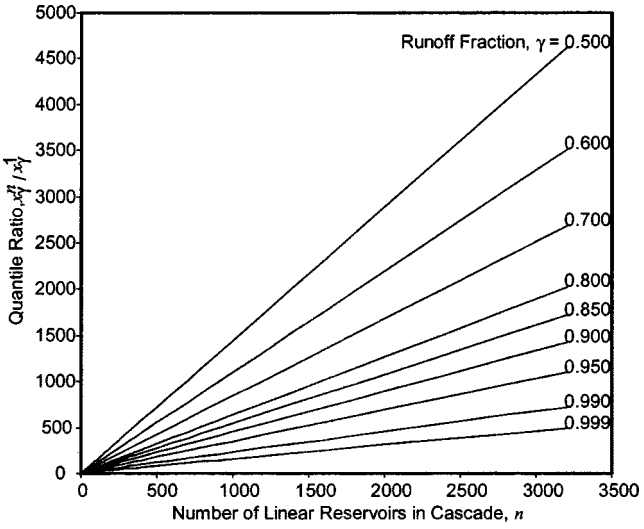

Fig. 2. Quantile ratios

Temporal Scaling

Mass balance computations associated with Eqs. (1)–(4), (7), (8), (12), and (13) are performed for each time interval (usually the daily interval), with end-of-interval storages becoming beginning-of-interval storages for the succeeding time interval. The time interval can be any length as far as the solution of the equations is concerned. Although the solution remains exact for any time interval length, the efficacy of the assumptions used in deriving the equations suffers with increasing length. That is, the assumptions of constant precipitation, potential evapotranspiration, and upstream surface flow over the time interval are poorer representations of reality as the time interval increases. Although the use of constant precipitation and potential evapotranspiration over the daily time interval in the lumped-parameter LBRM has proved adequate in the past, the additional use of constant upstream surface flow over the time interval in the distributed-parameter LBRM must yet be assessed.

Additionally, temporal scaling issues are associated with the spatial scale; it may be more appropriate to use finer time scales (small time intervals) with fine spatial scales. This can be assessed in application but is limited by the availability of meteorological and hydrologic data.

Flow Network

We consider that a watershed is broken into a group of cells, as in the map of Fig. 3. Each cell has flow properties assigned to it and one of eight flow directions, based on the watershed topography. Each cell has runoff from its surface and subsurface components into its surface channel system, and it has flow from an upstream cell into its surface channel system (except for the most-upstream cells). There are several general requirements for watershed maps such as Fig. 3. One and only one outlet from the watershed must exist. In the cells in Fig. 3, there must be one and only one cell in the watershed whose flow enters an "empty" cell (a cell with no flow designated in the watershed, i.e., a cell that is not within the watershed). All other cells with flows must enter other cells with flows, that is, other cells within the watershed. Furthermore, no "flow loops" may exist, isolating cells from drainage to the outlet. Fig. 4 shows two flow loops.

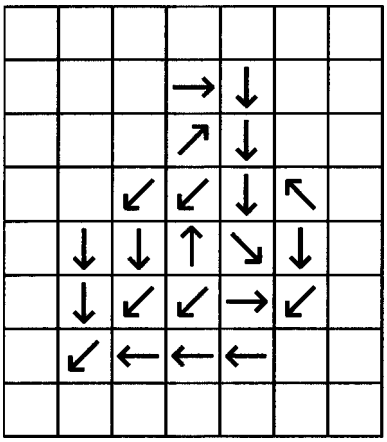

Fig. 3. Watershed grid flows

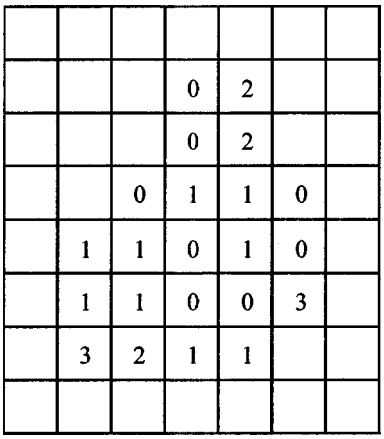

Fig. 5. Node codes

Cell Organization

Information on flow directions can be used to organize runoff and routing computations. By assigning each cell a number, subsequent routing computations may be made according to these numbers. Proper selection of the numbers enables efficient computations, whereby each cell's runoff and upstream flow are determined and routed through the cell only once, with minimum storage of pending hydrographs. A flow hydrograph out of a cell must be saved as a "pending" inflow hydrograph into the next downstream cell, until all upstream inflows for that next cell are computed; then they are added together to determine the total upstream surface flow into that next cell. Croley (1980) presented a microhydrology computation-ordering algorithm for application to a well-defined stream network to order subwatershed hydrograph computations and stream channel routing computations. It is the basis for a computation routing algorithm for watershed cells, as shown in Fig. 3. Each cell in a watershed map, such as Fig. 3, is regarded as a node in a flow network. Each node must be coded as to the number of flows that exist into it (0 through 7). The example of Fig. 3 is coded in Fig. 5. To avoid manual ordering of flow computations for very large networks, the rules for numbering nodes are explicitly stated in the logic flowchart of Fig. 6 and enable efficient hydrograph modeling and subsequent routing from cell to cell (node to node). The example of Figs. 3 and 5 is illustrated in Fig. 7.

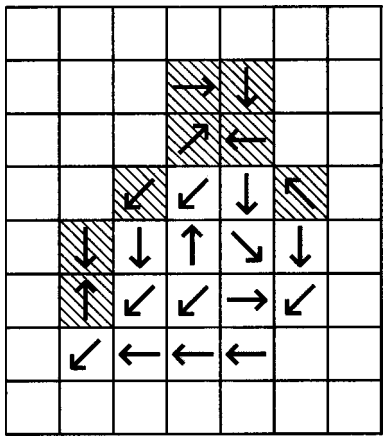

Fig. 4. Isolated flow paths

In Fig. 6, two broad arrows, labeled A and B, represent auxiliary storage operations for computer implementation of the algorithm. Arrow A represents a "push" onto a last-in, first-out storage stack, at each node traversed going upstream, of the branch direction taken out of the node. It keeps track of the directions taken at each node and is useful in determining the branch directions that are left to investigate later. Arrow B represents a "pop" from the stack, at each node traversed going downstream, of the branch direction last taken upstream from the node. It determines the next upstream branch to investigate from this node.

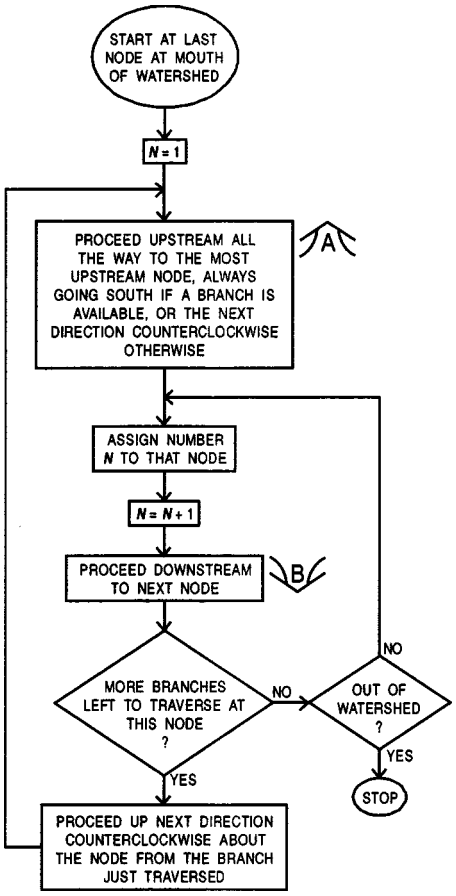

Fig. 6. Node numbering algorithm

			3	5		
			4	6		
		19	16	7	2	
	20	17	15	8	1	
	21	18	13	9	10	
	22	14	12	11		

Fig. 7. Node numbers

By comparing the map of node numbers with the original map of flow directions, flow loops or other isolated flow paths may be identified. Of course, for routing computations to proceed, there should be no isolated flow paths. As an example of this identification, the flow directions map of Fig. 4 is used with the flow algorithm of Fig. 6, resulting in the node numbers map in Fig. 8. The groups of cells within the two isolated flow paths shown in Fig. 4 have no node numbers assigned to them.

Flow Routing

The procedure for computing hydrographs for each cell and routing them downstream according to the numbering and coding scheme is detailed in the logic flowchart of Fig. 9. Again, a last-in, first-out stack is used for auxiliary storage; this time it efficiently saves the pending hydrographs (arrow A) and releases them as needed (arrow B), minimizing computer storage requirements while modeling each cell's hydrology only once.

Recursive Routing

The routing computations are actually programmed with a recursive routine, wherein the routine for determining the flow out of a cell involves successively calling itself to determine the flow out of other cells entering the cell. This implementation does not require the use of cell codes or numbers; it only requires that the watershed outlet cell be known for the first call to the routine. However, cell codes and numbers are calculated, since they are useful in checking the flow map for proper definition (outlet cell exists, no flow loops) and determining the outlet cell location before using it in the routing computations.

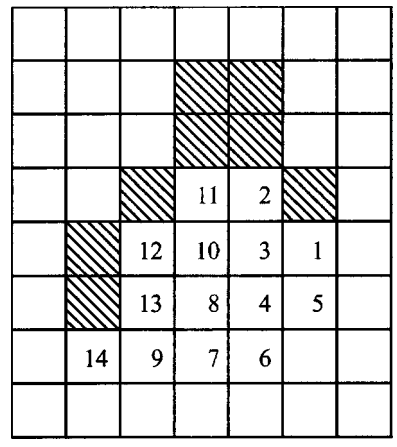

Fig. 8. Node numbers for map of Fig. 2

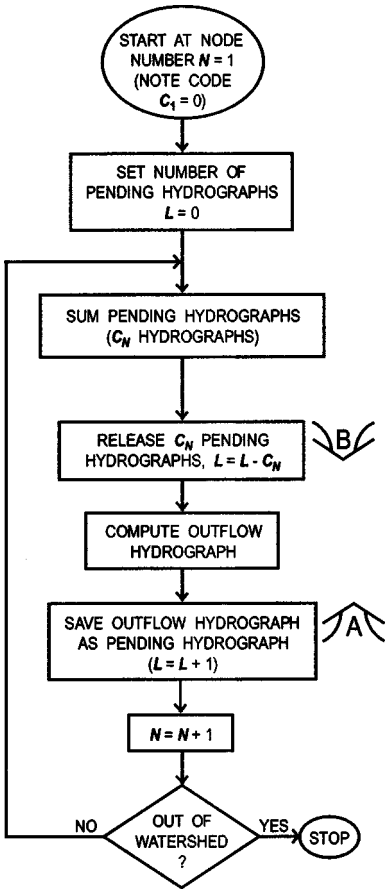

Fig. 9. Routing algorithm

Computer Implementation

The distributed-parameter LBRM is implemented as a routing of the outputs from the lumped-parameter LBRM, as described here, applied to each grid cell. This arrangement requires a fairly substantial amount of computing resources. For example, the Kalamazoo River watershed in southwest Michigan has 5,612 1-km^2 cells, and a 10-year model simulation requires computer storage for five daily water volumes (see Fig. 1) in each cell in the amount of 391 MB (5,612 cells $\times$ 3,652 days $\times$ 5 volumes $\times$ 4 bytes per value = 409,900,480 bytes). (Note that 1 MB = 1,048,576 bytes.) Likewise, three to five daily input meteorology variables correspondingly require 235 to 391 MB, daily runoff output requires 78 MB, an optional 11 daily internal flow rates require 860 MB, and 9 to 15 physical watershed characteristics data require 704 to 1,173 MB. The total requirement is between 1,407 and 2,893 MB. Since we use simulations that are longer than 10 years, we cannot store everything in memory for the entire simulation. Our first approach processed each cell separately for the entire simulation period. We entered meteorological station data and calculated daily input meteorology for a cell. Then we modeled the cell hydrology with the LBRM, retaining

only daily runoff output for the cell in memory, but reading input meteorology and physical cell characteristics from disk as needed and writing water volumes and internal flow rates to disk as needed. This approach minimized computer memory requirements to a single cell at any time (about 0.2 to 0.4 MB for the 10-year Kalamazoo River simulation example) but it required excessive input and output operations from and to hard disk, resulting in excessive computer time and disk wear. In addition, we were reading the same meteorological station data over and over.

We now process one year at a time for all cells, keeping all variables in memory until the end of the year. For each year, we read all the meteorological station data once into memory. Two hundred stations require only about 1 MB. This method implies reading the station data file once per year rather than once per cell. For a 50-year simulation of the Kalamazoo River watershed, we read the meteorological station data file 50 times instead of 5,612 times. Then we model the cell hydrology with the LBRM for each cell, routing flows through cells and keeping all variables in memory. For the Kalamazoo River basin, memory requirements are 1/10 of those cited in the last example, or 141 to 289 MB. At the end of each simulated year, the outputs for all cells are written to disk, thereby also significantly reducing input and output operations from and to hard disk and greatly improving performance. Computer time was reduced by greater than an order of magnitude. All memory is dynamically sized on the basis of the number of cells in the watershed, so memory requirements are scaled with the size of the watershed.

Summary

We briefly review microscale and macroscale watershed hydrology models and observe model requirements and limitations of both. Since macroscale operational models link the research community with policy and decision support institutions, we identify issues associated with applying a macroscale model to the microscale by looking at a specific case study. We modify GLERL's LBRM continuity equations to allow upstream inflow when the model is applied to a single cell within a watershed. We find the modifications in terms of corrector equations to be applied to the original equation solution. We also change the LBRM to use independent actual and potential evapotranspiration (appropriate for microscale use) instead of complementary actual and potential (appropriate for the macroscale). We find that linear reservoir coefficients for moisture storages of the upper and lower soil zones and the groundwater zone can be used across scales in initial approximations. However, initial approximations of surface storage coefficients for a network of cells are related to a macroscale coefficient through consideration of a cascade of linear reservoirs and characteristic travel times.

We organize LBRM applications to constituent watershed cells in a flow network by identifying the network flow cascade and then automatically arranging the cell computations accordingly. We identify required characteristics of any flow network map and design system checks to guarantee them. These characteristics include the presence of a unique watershed outlet cell and the absence of flow loops within the watershed. We devise a network cell numbering and coding scheme for these checks and for subsequently ordering LBRM computations and routing flows throughout the watershed. Finally, we outline the computer implementation for the microscale distributed LBRM and refine the implementation by application to the 5,612 1-km^2 cells of the Kalamazoo River watershed.

Acknowledgments

Partial support for Chansheng He came from the National Research Council Research Associateship Program, and Western Michigan University Department of Geography Lucia Harrison Endowment Fund is acknowledged while he was on his sabbatical leave from GLERL.

Notation

The following symbols are used in this paper:

b_1, b_2 = empirical coefficients relating solar insolation at the surface to cloud cover;
C = upper soil-zone moisture capacity;
$\mathbb{C}$ = units and proportionality coefficient;
d = time in one day;
E_g = daily evapotranspiration rate from the groundwater zone;
E_ℓ = daily evapotranspiration rate from the lower soil zone;
E_s = daily evapotranspiration rate from the surface zone;
E_u = daily evapotranspiration rate from the upper soil zone;
e_p = potential evapotranspiration rate;
G = groundwater-zone moisture content (subscript, when used, denotes time);
$g(t)$ = sum of time-dependent inflows to moisture storage Z;
H = nonlatent heat released to the atmosphere during the day;
h = surface flow input rate from upstream cells;
L = lower soil-zone moisture content (subscript, when used, denotes time);
$\overline{L}$ = average water volume in the lower soil zone during the day;
m_d = daily snowmelt on day d;
n = number of linear reservoir moisture storages in sequential cascade representing watershed;
Q = volume of basin outflow (from the surface storage) over time increment $(0, t)$;
Q' = volume of basin outflow (from the surface storage) over time increment $(0, t)$ for nonzero upstream surface flow input;
$Q(t)$ = accumulated outflow volume from bottommost linear reservoir in a cascade at time t;
$Q(x)$ = cumulative distribution function for the gamma distribution evaluated at x;
q_i = outflow rate from ith linear reservoir moisture storage in cascade;
$q(t)$ = outflow rate from bottommost linear reservoir in a cascade as a function of time t;
S = surface-zone moisture content (subscript, when used, denotes time);
S' = surface-zone moisture content for nonzero upstream surface flow input (subscript, when used, denotes time);
s = net supply rate;
T = daily average air temperature;
T_b = "base" temperature (a parameter to be determined in calibration);
$T_{\max}$ = maximum daily air temperature;
$T_{\min}$ = minimum daily air temperature;

t = time;
t_γ = time when the fraction of runoff, γ, occurs from a cascade of linear reservoirs (superscript, when used, denotes number of linear reservoirs in cascade);
U = upper soil-zone moisture content (subscript, when used, denotes time);
$\overline{U}$ = average water volume in the upper soil zone during the day;
V_g = groundwater flow volume into surface storage during time interval $(0,t)$;
V_i = interflow volume into surface storage during time interval $(0,t)$;
V_r = surface runoff volume into surface storage during time interval $(0,t)$;
x_γ = γ-quantile from the gamma distribution;
Z = moisture storage;
Z_i = moisture storage in ith linear reservoir in cascade;
α = linear reservoir proportionality factor;
α_d = linear reservoir coefficient for deep percolation from the lower soil zone to the groundwater zone;
α_g = linear reservoir coefficient for groundwater flow to the surface;
α_i = linear reservoir coefficient for interflow from the lower soil zone to the surface;
α_p = linear reservoir coefficient for percolation to the lower soil zone from the upper;
α_s = linear reservoir coefficient for outflow from the surface zone;
β = partial linear reservoir coefficients associated with evapotranspiration;
β_g = partial linear reservoir coefficient for groundwater-zone evapotranspiration;
β_ℓ = partial linear reservoir coefficient for lower soil-zone evapotranspiration;
β_s = partial linear reservoir coefficient for surface-zone evaporation;
β_u = partial linear reservoir coefficient for upper soil-zone evapotranspiration;
$\Gamma(n)$ = gamma function $(=\int_0^\infty e^{-x}x^{n-1}dx)$;
γ = fraction of total runoff volume;
γ_f = latent heat of fusion;
γ_v = latent heat of vaporization;
ρ_w = density of water;
Φ_d = daily surface solar insolation on day d;
Φ_s = cloudless daily solar insolation at the surface;
Ψ = total heat available for evapotranspiration during the day (subscript, when used, denotes the day); and
ω = linear reservoir coefficient for ith linear reservoir moisture storage in cascade.

References

Abdulla, F. A., Lettenmaier, D. P., Wood, E. F., and Smith, J. A. (1996). "Application of a macroscale hydrologic model to estimate the water balance of the Arkansas–Red River Basin." *J. Geophys. Res.*, 101(D3), 7449–7459.

Beven, K. J. (2000). *Rainfall-runoff modeling: The primer*, Wiley, New York.

Bicknell, B. R., Imhoff, J. C., Kittle, J., Donigian, A. S., and Johansen, R. C. (1996). *Hydrological simulation program—FORTRAN, user's manual for release 11*, U.S. Environmental Protection Agency, Environmental Research Laboratory, Athens, Ga.

Chow, V. T. (1964). *Handbook of applied hydrology*, McGraw-Hill, New York, 14-28 and 14-29.

Crawford, N. H., and Linsley, R. K. (1966). "Digital simulation in hydrology: Stanford Watershed Model IV." *Technical Report 39*, Dept. of Civil Engineering, Stanford Univ., Calif.

Croley, T. E., II. (1980). "A micro-hydrology computation ordering algorithm." *J. Hydrol.*, 48(3/4), 221–236.

Croley, T. E., II. (1982). "Great Lakes basins runoff modeling." *NOAA Technical Memorandum ERLGLERL-39*, National Technical Information Service, Springfield, Va.

Croley, T. E., II. (2002). "Large basin runoff model." *Mathematical models in watershed hydrology*, V. Singh, D. Frevert, and S. Meyer, eds., Water Resources Publications, Littleton, Colo., 717–770.

Croley, T. E., II, He, C., and Lee, D. H. (2005). "Distributed-parameter large basin runoff model. II: Application." *J. Hydrologic Eng.*, 10(3), xxx-xxx.

Engman, E. T., and Gurney, R. J. (1991). *Remote sensing in hydrology*, Chapman and Hall, New York.

Gray, D. M. (1970). "Energy, evaporation and evapotranspiration." *Handbook on the principles of hydrology*, D. Gray, ed., Water Information Center, Port Washington, N.Y., 3.1–3.65.

Koren, V. I., Finnerty, B. D., Schaake, J. C., Smith, M. B., Seo, D.-J., and Duan, Q.-Y. (1999). "Scale dependencies of hydrologic models to spatial variability of precipitation." *J. Hydrol.* 217, 285–302.

Leavesley, G. H., and Stannard, L. G. (1995). "The precipitation-runoff modeling system—PRMS." *Computer models of watershed hydrology*, V. Singh, ed., Water Resource Publications, Littleton, Colo., 281–310.

Liang, X., Wood, E. F., and Lettenmaier, D. P. (1996). "Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification." *Glob. Planet. Change*, 13, 195–206.

Martinez, J. E., Dunchon, C. E., and Crosson, W. L. (2001). "Effect of the number of soil layers on a modeled surface water budget." *Water Resour. Res.*, 37(2), 367–377.

Mohseni, O., and Stefan, H. G. (1998). "A monthly streamflow model." *Water Resour. Res.*, 34(5), 1287–1298.

Nash, J. E. (1957). "The form of the instantaneous unit hydrograph." *International Association of Scientific Hydrology, Publication 45*, 3, 114–121.

Quinn, P., Beven, K., and Culf, A. (1995). "The introduction of macroscale hydrological complexity into land surface-atmosphere transfer models and the effect on planetary boundary layer development." *J. Hydrol.*, 166, 421–444.

Rainville, E. D. (1964). *Elementary differential equations*, 3rd Ed., MacMillan, New York, 36–39.

Richardson, C. W., and Wright, D. A. (1984). "WGEN: A model for generating daily weather variables." *ARS-8*, U.S. Dept. of Agriculture, Agricultural Research Service.

Valeo, C., and Moin, S. M. A. (2001). "Hortonian and variable source area modeling in urbanizing basins." *J. Hydrologic Eng.*, 6(4), 328–335.

Zhao, R. J., Zhuang, Y.-L., Fang, L.-R., Liu, X.-R., and Zhang, Q.-S. (1980). "The Xinanjiang model." *Hydrological forecasting*, IAHS Publication No. 129, Wallingford, U.K., 351–356.

Zhu, A. X., and Mackay, D. S. (2001). "Effects of spatial detail of soil information on watershed modeling." *J. Hydrol.*, 248, 54–77.